Acoustic Properties of Consonant Sequences in Conversational French Speech

Danielle Duez

Laboratoire Parole et Langage, CNRS UMR 6057, Université de Provence

ABSTRACT

The current paper analyses spectrographic data of two-consonant sequences with one or both consonants identified, omitted or changed into another consonant, with one or two features identified or with no feature identified, when compared to an earlier perception analysis. Perceptual and acoustic data mostly correspond, proving that consonants had indeed been changed, significantly reduced or deleted, mainly in coda position and in non-final phrase syllables, thereby preserving acoustic information crucial for lexical access, integration of prosodic structure and successful communication.

1. INTRODUCTION

The present study consists of a spectrographic investigation of two-consonant sequences in spontaneous French speech, both consonants occurring initially or finally (/C1C2/) in a syllable and consonants occurring in two consecutive syllables (/C1. C2/ and /C1(ə).C1/). Spectrographic data were compared to underlying phonological sequences and perceptual data previously obtained [1].

Sections 1 and 2 respectively summarise spectrographic data on consonant sequences identified as the underlying phonological sequences and as different sequences. Section 3 examines the interactive effects of reduction and assimilatory processes by word class (function or content), position in syllables (onset or coda), and phrases (final or non-final).

2. SPECTROGRAPHIC ANALYSIS

2.1.CONSONANT SEQUENCE CHARACTERISTICS

In [1] 720 consonant sequences (CS's) extracted with adjacent vowels from conversations of two male French speakers (Spk1 and Spk2) were presented for identification to sixteen listeners. Two series of 360 $V_1C_1C_2V_2$'s (one per speaker) contained nine 40 C_1C_2 sequence groups: OO, OF, OS, FF, FO, FS, SS, SO, SF, where O is an occlusive, F a fricative and S a sonorant. Each CS was assigned a single identification value corresponding to the phonological underlying sequence reported by at least 75% of listeners. This yielded two main groups:(1) CS's identified as the underlying phonological CS's and (2) CS's identified as different (Spk1: 179, Spk2: 166). The latter has four subdivisions: (2a) the phonological C1 and/or C2 was identified as another consonant (C), (2b) only one or two features were identified, (2c) the C was omitted, and (2d) neither a C nor a feature was identified.

2.2. METHOD

Segmentation of C's. For each CS spectrograms and oscillograms were generated. CS onset coincides with the preceding-V offset and CS offset with following-V onset. C1/C2 boundaries follows:

O+(O, F or S): When there is a burst, the boundary is always at the end of the burst of the first O (voiced or unvoiced). In absence of a burst, the boundary is the discontinuity (if visible) between the two occlusions (O+O), between occlusion end and noise beginning (O+F), or the beginning of S-F*2* (O+S).

F+(F, O or S): The F/F-boundary location is based on changes in spectral zone, discontinuity in voicing, and change in intensity. The complete disappearance of F noise coincides with the beginning of occlusion (F+O) or formants (F+S).

S+(O,F or S). The cessation of the second formant coincides with the apparition of voice bar or silence (S+O), of frication noise (S+F) or the change in harmonic density, spectral repartition of formants (S+S).

Acoustic-cue analysis. Acoustic patterns of C's were examined in relation to perceived and phonological C's.

O's [p, t, k, b, d, g]. The presence of occlusion and/or burst was controlled, when visible their duration and burst frequency were measured. The presence of low frequencies was checked for voiced O's.

F's [f, s, ʃ, v, z, ʒ]. The presence of noise was controlled, its frequency limit and maximum measured. Voiced F's may have two or three formants, depending on the proportion of voicing and fricative elements. Voiced-F patterns were examined and visible formants measured.

S's [m, n, l, ʀ]. Formants and duration were measured. It is known that the spectrum of nasals is dominated by a strong low-frequency formant and has mid-frequency formants. Laterals are characterized by stronger mid-frequency formants than nasals. The French /ʀ/ is highly context dependent (with formants or noise).

Reduction and assimilation analysis. For oral O's reported as nasals, the presence of mid-frequency formants was examined. Three patterns of nasalisation (N) were determined: no N (no visible formants), partial N (a separate stop exhibiting an occlusion and/or burst) and total N (no interruption of mid-frequency formants).

Similarly, three patterns of voicing or devoicing were defined for O's and F's, depending on the presence of low frequencies. The same procedure was used to investigate frication of O's. For place change of O's and F's, burst frequency, noise limits and maximum were measured, for nasal-place change V/C transitions were studied. Identified and unidentified features were investigated as a function of the above criteria. The absence of noise, closure and burst, and formants was checked for omitted F's, O's and S's, respectively. The same was done for C's with no feature identified. The duration of remaining C1's and C2's was compared with the same intervocalic consonants (IC's).

Syllabic structure, position in syllable, word type and prominence. CS's were divided into homosyllabic [. C_1C_2 or C_1C_2.] and heterosyllabic [C_1 . C_2] or C1. C1]. In [C_1 . C_2]'s, C1 and C2 are coda and onset C, respectively. [C1. C1]'s result from the deletion of an optional mute /ə/ within a word or in a function word [C1(ə). C1]. C's were also analysed as a function of location in function or content words. With no lexical stress in French, prominent are mostly phrase-final syllables to which prominence was assigned.

3. C'S IDENTIFIED AS PHONOLOGICAL

3.1. ACOUSTIC CHARACTERISTICS OF O'S

O's with a burst outnumber those without, confirming the role of burst in O's identification. For example, Spk2 has 44 C1's and 90 C2's with a burst; 19 C1's and 14 C2's without a burst. O's with a burst are at least 15% longer. Burst duration lengthens as the place of articulation moves backward (e.g. C2- /p, t and k/ is 20, 32 and 40 ms long). Burst frequency varies significantly with place of articulation (Spk1, C1, F(5,26)=7; C2, F(4,)=9; Spk2, C1, F(5), C2, F(10), p=0.0001, and within a category, especially for dentals (a range of about 3000 Hz), reflecting context effects. For example, there is complete overlap of noise and occlusion for one /t/ and 1/d/, and partial overlap for 2/t/'s, all located in a fricative context. Effects of context are also visible for 9 /O/'s preceded by a nasal vowel. All have complete overlap of mid-frequency formants and occlusion, seven also have a burst.

3.2. ACOUSTIC CHARACTERISTICS OF F'S

/f/-mean limits are very low (Spk1: C1: 1186 Hz ; C2: 1621 Hz ; Spk2: C1: 1006 Hz, C2: 1220 Hz) and noise is very weak, often invisible. Limit and maximum values of /ʃ/'s and /ʒ/'s are very similar (C1. /ʃ/: 2694 Hz(Spk1), 4016 Hz (Spk2); /ʒ/: 2725 Hz (Spk1) and 3725 Hz (Spk2). Mean maximum values of /s/'s and /z/'s range from 4416 Hz to 5000 Hz and limits from 2175 Hz to 2996 Hz. Again, F's have visible effects of context, such as high variability of dentals due mainly to labialisation, and the change of some /v/'s into approximants (5 C1's and 4 C2's have formants, 1 C2 has both formants and noise and 1 C2 noise) and the partial devoicing of some voiced F's.

3.3. ACOUSTIC CHARACTERISTICS OF S'S

About 80% of /l, m, n and R/'s have mid-frequency formants. /l/ values are around 1600 Hz, 2500 Hz and 3500 Hz; those of /m/ and /R/ around 1250Hz and 2500Hz. Context effects are visible, especially for /ʀ/s. For example, for C2's, 14 /ʀ/s have mid-frequency formants, 5/ʀ/s followed or preceded by a /t/ or an /f/ have a complete (5) or partial (1) unvoiced fricative spectrum with frequency limits around 1000 Hz; 1/ʀ/ has no low frequencies but mid-frequency formants, 1 has both voicing and noise, 2 are changed into vowels, mostly in coda position, for 5 there is no information.

4. C'S IDENTIFIED AS DIFFERENT

4.1. CONSONANTS REPORTED AS ANOTHER CONSONANT

Oral O's reported as nasals. Eight O's (6/d/'s and 2/t/'s) preceded by a nasal vowel and reported as /n/'s have complete overlapping of at least two mid-frequency formants with occlusion, and no burst. The /d/'s are 45 ms, 57 ms, 64 ms, 67 ms and 77 ms long, the /t/ are 94 ms and 55ms. There is also a totally nasalised /p/ followed by /ɔ̃/ and reported as /m/ (74 ms long and has a 1000-Hz formant). Two /b/'s, 1/d/, and 1 /t/ reported as nasals are in an oral V context: these may be confusions. One of the 2 /b/'s has formants (1100 Hz and 2600 Hz) and is 67 ms long. The /d/, the /t/ and the other /b/ have no formants.

Voiced and unvoiced C's reported as unvoiced and voiced

C1's. No low frequencies for 4/ʒ/'s, 1/v/ and 1/z/ and 4/d/'s reported as unvoiced; on the contrary, there are low frequencies for 4/s/, 1/p/ and 1/t/ reported as voiced. In all cases, there is anticipatory effect of a unvoiced or voiced C2. Although reported as unvoiced 1/ʒ/ is partially voiced (25 ms out of 50ms), 1/ʒ/, 1/v/ and 1/d/ have low frequencies. These may be confusions induced by the excising technique used in perceptual tasks.

C2's. No low frequencies, but noise for a 51-ms /v/ reported as /f/ in /sv/ .

C1+C2's. /ts/ reported as voiced has low frequencies, /dʒ/ reported as unvoiced has no voicing: these may be misproductions and/or misperceptions. There are 6 single C's resulting from the coalescence of C1-manner-and-place and C2-voicing: /sd/=>/z/, /sv/=>/z/, /fz/=>v/, 2/ʒ/s=/ʃ/, /tv/=>/d/. The /z/'s and the /v/ have low frequencies, their respective duration is 150 ms, 120 ms, and 94 ms. The 120-ms /d/ has a voice bar and burst (3750 Hz). The two /ʃ/'s have no low frequencies and a duration of 100ms and 175 ms. Each C is longer than IC's.

Other cases

C1's. Two /n/'s reported as /m are preceded by a labial V. One has the clear labial-downward movement at the end of preceding V and three formants (1117 Hz, 2194 Hz and 4209 Hz), it is 65 ms long. The other is 20 ms long, having

a 2500-Hz formant and flat V/C transitions.. A /f/ results from /t/-manner-and-place change, another from a /s/-place change (/st/=>/ft/). Both have a weak noise (limit around 1000 Hz), they are 60 ms and 115 ms long. A /ʀ/ reported as /b is followed by /l/. With no occlusion, burst, or formants, this may be a confusion.

C2's. There is carryover effect of an /s / for a /p/ changed into an /f/ with a weak frication noise (limit: 500 Hz, duration: 65 ms). A /l/ preceded by a /s / is changed into a typical /t/ (duration: 64 ms; burst: 4200 Hz). A /b/ reported as /g/ has a voiced bar (62ms), and a burst (18ms, frequency: 1669Hz). A 105-ms /d/ reported a /v/ has a 1200-Hz formant. The /t/, /g/ and /v/ may be errors.

C1 and C2. A typical /z/ (limit: 3500 Hz; max, 4500 Hz) results from the merger of /d/ place and /v/ manner.

4.2. C'S WITH ONE OR TWO IDENTIFIED FEATURES

C1. There is partial or complete carryover effect of a preceding nasal V for 2 /d/'s and one /t/ reported as oral or nasal and 3 /d/'s reported as /n or m/; also no formants for a /d/ reported as /d and n/. Complete N is visible for 5/t/'s, 1/d/ and 1/p/, partial N for 1/t/ and 2/d/'s associated to complete voicing, partial voicing (1/t/) and devoicing (3/d/'s). For O's reported as nasals, voiced or unvoiced, there is both effect of the preceding nasal and C2 voicing. Three /k/'s reported as /k or g/ are totally voiced and nasalised. There are no mid-frequency formants for 4 /b/'s, reported as /p, b and v/ and 1/p/ reported as /b, p and m/, one 1000-Hz formant for 1/b/ reported as /b, m, v and w/. Out of nasal context, these may be confusions.

There is complete anticipatory effect of a voiced C2 for 9 unvoiced F's and 3 unvoiced O's, of an unvoiced C2 for 1 voiced O and 1 voiced F, partial devoicing for 4 voiced F's and partial voicing for 3 unvoiced F's and 4 unvoiced O's. However, 16 O's and F's reported as devoiced or voiced, are followed by a C2 with same voicing. This confusion may be due to the artificial truncation of V's.

Place of articulation was not clearly identified for 5/n/'s and 2/m/'s and 2/l/'s partly because of the absence of clear V/C transitions. One /d/ reported as /l, n,m/ is changed into an approximant (with a 2000-Hz formant). Listeners could not identify the place of 3 O's and 3 F's, partially assimilated to C2. For example, 1 /p/ identified as /p and t/ has a 3000-Hz burst; 1 /f/ identified as /ʃ and f/ has /f/-limit (1500Hz) and /ʃ/ maximum (3000 Hz). All C1's are 30% shorter than their identified counterparts; this may prevent identification.

C2. C2 voicing effect is visible for 2 /s/'s and 1/f/ preceded by a voiced C1: the /s/'s has partial voicing, the /f/ noise and formants. Two /t/'s are partially voiced, 1/ʒ/ partially devoiced although having the same voicing feature as C1. No effect of C1-voicing feature for 2/t/'s, 1/p/, and 4/f/'s and 2/d/'s reported as voiced and unvoiced: these may be confusions. Four O's preceded by an F and one /t/ followed by /i/ were partly reported as their F counterparts: there is partial frication for three of them, voicing for 1/p/ and formants for 1/d/. Listeners could not clearly identify the place of 1/l/ and 1/m/, 2 /s/'s, 1/f/ and 1/z/: /l/ and /m/ have no clear transitions, the /s/'s preceded by /y/ and /ɔ̃/, and the /z/ followed by /f/ are partially labialised, there is partial place assimilation (with /f/ limits and /s/ max) for an /f/ preceded by /d/. A short /d/ (42ms) with an approximant spectrum (FM: 1381Hz and 2321 Hz) is reported as /d, v and w/.

C1 and C2. Effects of vowel context and/or reciprocal effect of both C1 and C2 can be observed for 7 of the 13 /C1C2/'s with C1 and C2 incompletely identified. For example, in a /tk/ sequence preceded by a nasal vowel and followed by a labial there is /t/ nasalisation and /k/ labialisation. The /t/ reported as /t, d and n/ is voiced and has 3 mid-frequency formants, the /k/ reported as /k, p and v/ has noise. There is complete voicing of /k/ and partial devoicing of /z/ in /kz/. However, there is no correspondence between acoustic patterns and responses for 5 /gz/'s. The /g/'s and /z/'s reported as voiced and unvoiced are completely voiced. It is the same for a /ʒd/ sequence: the /ʒ, perceived as /ʒ/ and /ʃ/, is voiced, the /d/ reported as /d/ and /l/ has no formants.

4.3. C'S WITH UNIDENTIFIED FEATURES

C1's. Most C1's were partially omitted, because of exceptional shortness and/or incomplete acoustic information. For example, 5/l/'s have mid-frequency formants, but their duration ranges from 25 ms to 53 ms. Eleven /ʀ/'s have formants, 4 are changed into vowels, 2 have noise, one is deleted. Three /m/'s have formants, the fourth is devoiced, all are short (from 27ms to 35 ms). O's and F's are exceptionally short, strongly reduced or omitted. For example, there is a short duration for 1 /t/ and 4 /d/'s (ranging from 35ms to 45 ms), an approximant spectrum for 2 of the /d/'s, 8 /k/'s and /v/'s and deletion for 1/p/ and 1/d/ and 2/k/'s. There is no possible measurement for 4 F's and 16 O's.

C2's. As for C1's, unidentified C2's are characterised by shortness, reduction and deletion. Two /l/'s have formants and are 20ms and 18 ms long, 2 have both formants and noise (their duration is 75 ms and 25 ms), 2 are devoiced and have noise, their duration is 53 ms and 26 ms, 1 is partially devoiced and 21 ms long, the last one is deleted. Five /ʀ/'s have formants and a duration ranging from 17 ms to 45 ms, 1 /m/ and 1 /n/ have formants, their duration is 45 ms and 50 ms, respectively, another /m/ is deleted. One /p/ with a 1200-Hz burst, reported as /m/ by 10 listeners, is followed by /ɔ̃/ and may be nasalised. One /k/ is completely voiced, the second an approximant. Two /v/'s are deleted, a third one is devoiced and 41 ms long. One short /s/ (35 ms) and one short /f/ (31 ms) are partially voiced. One /f/ and 1/d/ could not be measured.

C1's and C2's. /dl/ has a fricated /d/ and a /l/ with 3 formants, /ʀd/ has a vowel-like /ʀ/, /vʀ/ has an approximant /v/ and a very short /ʀ/ (18 ms), /zs/ has a partially voiced /s/, /vs/ a labialised /s/. /zs/ and /vs/ are long (90 ms and 153 ms).

4.4. OMITTED C1'S AND/OR C2'S

C1's. Omitted C's are mostly S's with no visible formants. Three exceptions (1/l/ and 2 /ʀ/'s with mid-frequency formants) are very short (30ms, 18 ms, 30 ms). The remaining C2's have a duration equal or inferior to IC's, which suggest a complete deletion of S's. Two identical C's usually coalesce into a single C whose duration exceeds that of IC's. (e.g. /ʀ/ duration: 96ms, 76 ms for Spk1 and 77 ms for Spk2: mean IC duration: 58 ms and 52 ms). C1's and C2's with the same place and manner also tend to combine into a single C whose duration is longer than that of IC's: Spk1 (/s/: 156 ms, IC /s/: 104 ms); Spk2 (//f/: 145 ms, IC //f/: 88 ms; /t/: 94 ms and 122 ms, IC /t/: 91 ms). For C1's and C2's with the same manner and voicing, the trend is similar although less marked. The remaining /z/, /p/ and /d/ in /gz/, /tp/ and /kd/ are much longer (100 ms, 114ms and 105 ms) than IC's (70 ms, 90 ms and 61ms). The long duration of most C's resulting from the fusion of C1 and C2 suggests an assimilation process rather than deletion.

C2's. One omitted /m/ and 2 /l/'s have formants. Their short duration (from 15ms to 20 ms) may explain listeners' omission. There are no formants for 3 omitted /ʀ/'s/ and 5 /l/'s. The duration of the remaining C1's does not exceed that of IC's suggesting the S's deletion. Two deleted /d/'s and 1/t/ are preceded by a /z/, and 2 deleted /t/'s belong to /st/ sequences. The remaining /z/'s have a duration (75 ms) close to that of IC-/z/'s (78 ms), contrary to the two /s/'s which are much longer (145 ms and 175 ms, respectively) than IC-/s/'s. The remaining /t/ of /td/ has no voice bar and a duration equal to that of IC's. The long duration of the remaining /s/'s again suggest an assimilation process.

5. SYLLABLE STRUCTURE, WORD TYPE, LOCATION IN SYLLABLES AND PHRASES

For Spk1 and Spk2, 137 and 122 out of the 151 and 142 C1's reported as different belong to heterosyllabic CS's. They are mainly coda C's (Spk1: 110, Spk2: 93) or the first C of a [C1. C1] sequence (Spk1: 27; Spk2: 29] and located in non-final phrase syllables. The effect of syllable structure, location within syllables and prominence on the repartition of consonants identified as phonological and different is significant for both Spk1 and Spk2, for Spk1 there is also significant effect of word type. The interactive effect of syllable structure, position in syllables and phrases is particularly powerful for C1's. For Spk1, 121 C1's reported as different are in non-final syllables, 99 of them are in coda position, 22 in C1. C1's. The trends are similar for Spk2: of 103 C1's in non-final syllables, 80 are coda C's, 23 the first C1 in C1. C1's.

6. CONCLUDING REMARKS

With little context, the acoustic information contained in $V_1C_1C_2V_2$'s enabled listeners to identify about 50% of CS's. The loss of information may be partly due to confusions, sometimes induced by the artificial truncation of vowels; or partly due to misproductions, frequent in spontaneous speech. However, perceptual and acoustic data are usually in agreement, proving that many C's are indeed changed, significantly reduced or deleted.

Two main trends emerge for assimilated C's: 1) nasalisation of O's resulting from the overlapping production of the nasal vowel and closure gestures, and 2) devoicing or voicing of voiced or unvoiced C1's or C2's due to the anticipatory and/or carryover effect of an unvoiced or voiced C1 or C2. Similarly, C's with one or two identified features are partially nasalised and/or partly voiced or devoiced. Strikingly, most of the partially or completely assimilated C's keep their place of articulation, confirming previous results on stops [2]. Omitted C's and C's with no identified features are mostly S's. Since S's have a formantic structure, they are easily changed into vowels or deleted, reflecting a strong decrease in the magnitude of movements associated with the gesture or complete deletion of the gesture [3].

Onset C's in final-phrase syllables are less assimilated and reduced than coda C's in non-final phrase syllables. Onset C's are produced with more articulatory strength than coda C's; also, they have a higher signalling value in word recognition. In French, final-phase syllables are subject to lengthening and boundary markers, which may limit hypoarticulation processes. The results highlight a compromise between gesture economy and the need for successful communication [4].

REFERENCES

[1]D. Duez,"Consonant sequences in Spontaneous French speech", *SpoSS*, ESCA, La Baume les Aix, pp.24-26, 1998.

[2]D. Duez, "On spontaneous French speech: aspects of the reduction and contextual assimilation of voiced plosives", *Journal of Phonetics*, 25. 407-427, 1995.

[3]C.P. Browman, and L. Goldstein, "Articulatory phonology: An overview", *Phonetica*, 49, 155-180, 1992.

[4] [1] B. Lindblom, "Explaining phonetic variation: a sketch of the H and H theory". In *Speech production and speech modelling* (W. Hardcastle and A. Marchal, editors), Vol 55, pp. 403-439. NATO ASI Series. Dordrecht, Boston and London: Kluwer Academic Publishers, 1990.